



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>7</sup> :</b> <b>C12N</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 00/34435</b> <b>(43) International Publication Date:</b> 15 June 2000 (15.06.00)
<b>(21) International Application Number:</b> PCT/US99/28451 <b>(22) International Filing Date:</b> 1 December 1999 (01.12.99) <b>(30) Priority Data:</b> 09/206,887 8 December 1998 (08.12.98) US <b>(71) Applicant:</b> CLONTECH LABORATORIES, INC. [US/US]; 1020 East Meadow Drive, Palo Alto, CA 94303 (US). <b>(72) Inventors:</b> LI, Xianqiang; 3989 Aiddlefield Road, Palo Alto, CA 94303 (US). ZHAO, Xiaoning; 110 E. Remington Drive #20, Sunnyvale, CA 94087 (US). JIANG, Xin; 1133 Rejcovich Way, San Jose, CA 95120 (US). FANG, Yu; 43059 Starr Street, Fremont, CA 94539 (US). DUONG, Tommy; 2491 Glen Elm Way, San Jose, CA 95148 (US). KAIN, Steve; 403 Read Avenue, Belmont, CA 94002 (US). <b>(74) Agent:</b> ADLER, Benjamin, A.; McGregor & Adler, 8011 Candle Lane, Houston, TX 77071 (US).		<b>(81) Designated States:</b> JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> <i>CIS</i> -ELEMENT REPORTER CONSTRUCTS AND USES THEREOF  <b>(57) Abstract</b>  The present invention provides <i>cis</i> element-reporter constructs for measuring transcription. A series of <i>cis</i> elements are linked to a reporter gene to generate <i>cis</i> -acting reporters. Examples of reporter genes are secreted alkaline phosphatase (SEAP), destabilized green fluorescent protein (d2EGFP) and luciferase. Also provided are various applications of such constructs.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

**CIS-ELEMENT REPORTER CONSTRUCTS AND USES THEREOF**

5

**BACKGROUND OF THE INVENTION**Cross-Reference to Related Applications

10 This application claims benefit of priority of pending non-provisional application U.S. Serial No. 09/206,887, filed December 8, 1998.

Field of the Invention

15 The present invention relates generally to the field of molecular biology of transcription. More specifically, the present invention relates to *cis* element reporter constructs for measuring transcription.

Description of the Related Art

20 Regulation of the activity of transcription factors is believed to be one of the most important biochemical controls to signal transduction, cell proliferation, differentiation, apoptosis, development, and disease. Studying the regulation of an individual transcription factor in various specific biological functional system has yielded a tremendous amount of information; however, the  
25 information that one can obtain from the current strategies of studying an individual factor is limited, despite the global functional status of a number of transcription factors. The information is limited because cross-talking among the different signal transduction pathways leads to activation of multiple transcription factors. Therefore, there is a need

to provide a tool to study the interaction and overall effect of various transcription factors.

Green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* is a reporter molecule for monitoring gene expression and protein localization *in vivo*, *in situ* and in real time (1-4). GFP fluoresces bright green upon mere exposure to UV or blue light-unlike other bioluminescent reporters which require additional proteins, substrates, or cofactors to emit light. GFP fluorescence is stable, species independent and can be monitored noninvasively in living cells. GFP fluorescence persists in formaldehyde-fixed cells and is well suited for double-labeling experiments with other fluorescent markers, including the GFP variant enhanced green fluorescent protein (EGFP, GFP<sub>mut1</sub>; 5-6).

EGFP encodes a protein which has a single, red-shifted excitation peak and fluoresces about 35 times more intensively than wild type GFP when excited at 488 nm (7), due to an increase in its extinction coefficient (Em). To ensure maximal mammalian expression, the coding region of EGFP contains more than 190 silent base mutations which correspond to human codon-usage preferences (7). The red-shifted spectrum and increased expression of EGFP make it ideal for fluorescence microscopy and fluorescence-activated cell sorting (FACS; 5, 8).

Destabilized EGFP (d2EGFP) contains the PEST domain from mouse ornithine decarboxylase (MODC), fused to the C-terminus of EGFP (9). This domain targets EGFP for rapid turnover effectively reducing EGFP's half-life to two hours. The introduction of d2EGFP greatly increases the utility of GFP in studying dynamic cellular-events *in vivo*.

Another reporter molecule, secreted alkaline phosphatase (SEAP, 10-11), is a secreted form of human placental alkaline

phosphatase and has been used as a reporter for the analysis of *cis*-acting DNA sequences and *trans*-acting factors. Secreted alkaline phosphatase is efficiently secreted from cells, and the levels of SEAP activity in the culture medium are directly proportional to changes in intracellular concentrations of SEAP mRNA and protein (10, 12). Secreted alkaline phosphatase differs from native placental alkaline phosphatase (AP) only in the absence of a membrane-anchoring domain. The secreted nature of secreted alkaline phosphatase offers many advantages over other transcription reporter molecules. Most notably, it is not necessary to prepare cell lysates to assay for the reporter protein. Because the transfected cells remain intact, the kinetics of gene expression can be studied easily using the same cultures by sampling the medium repeatedly; therefore, set-up and variability for multiple cultures are essentially eliminated.

Another important advantage of secreted alkaline phosphatase as a transcription reporter is that background signals due to endogenous alkaline phosphatases are nearly absent. Unlike most endogenous alkaline phosphatases, secreted alkaline phosphatase is extremely heat-stable and resistant to the inhibitor L-homoarginine. Therefore, endogenous alkaline phosphatase activity can be eliminated effectively by preheating the culture medium sample as 65°C and assaying in the presence of L-homoarginine.

In addition, luciferase is a widely used reporter. It is one of most sensitive reporters and like GFP and SEAP, may be fused to a transcription factor of interest and used to track the activation of the transcription factor of interest.

The prior art is deficient in the lack of effective means of detecting and measuring the activities of a number of transcription factors. Further, the prior art is deficient in the lack of a construct for

measuring transcription. The present invention fulfills this long-standing need and desire in the art.

5

## SUMMARY OF THE INVENTION

The present invention is directed to novel *cis*-acting reporters for measuring transcription. In one embodiment of the present invention, there is provided a *cis* element-reporter construct comprising a *cis* element, a reporter gene and a promoter. Representative *cis* elements include nuclear factor- $\kappa$ B, heat responsive element, Myc, p53, nuclear factor of activated T cells, activator protein 1 (AP1), serum responsive element, glucocorticoid responsive element and cAMP responsive element. Representative reporter genes include secreted alkaline phosphatase (SEAP), destabilized green fluorescent protein (d2EGFP) and luciferase (Luc). Representative promoters include thymidine kinase promoter and gonadotropin  $\alpha$ -gene promoter.

In another embodiment of the present invention, there is provided a method of generating a *cis* element-reporter construct by linking a *cis* element to a promoter and then fusing the element to a reporter gene.

In yet another embodiment of the present invention, there is provided a method of monitoring activation of a transcription factor, comprising the steps of generating a vector comprising a *cis* element, a reporter gene and a promoter, wherein the *cis* element is a component of a DNA sequence to which said transcription factor binds to ensure activity; transfecting a cell line with the vector; and detecting expression of the reporter gene, wherein expression of the reporter

gene indicates activation of the transcription factor. Representative cell lines include CHO, HEK 293, Saos2 and HeLa.

5 In still yet another embodiment of the present invention, there is provided a method of determining whether a gene is involved in a signal transduction pathway, comprising the steps of generating a vector comprising a *cis* element, a reporter gene and a promoter, wherein the *cis* element is a component of a DNA sequence which affects the expression of the a gene; expressing the vector in a cell line, wherein the cell line is associated with the signal transduction pathway; 10 and detecting the expression of the reporter gene, wherein expression of the reporter gene indicates the involvement of the gene in the signal transduction pathway.

Other and further aspects, features, and advantages of the present invention will be apparent from the following description of 15 the presently preferred embodiments of the invention given for the purpose of disclosure.

## BRIEF DESCRIPTION OF THE DRAWINGS

20

So that the matter in which the above-recited features, advantages and objects of the invention, as well as others which will become clear, are attained and can be understood in detail, more particular descriptions of the invention briefly summarized above may 25 be had by reference to certain embodiments thereof which are illustrated in the appended drawings. These drawings form a part of the specification. It is to be noted, however, that the appended drawings illustrate preferred embodiments of the invention and therefore are not to be considered limiting in their scope.

30

**Figure 1** shows the restriction map of pNF- $\kappa$ B-SEAP.

**Figure 2** shows the restriction map of pNF- $\kappa$ B-d2EGFP.

**Figure 3** shows the restriction map of pNF- $\kappa$ B-Luc.

**Figure 4** shows mediated induction of transcription factors in HEK 293 cells. Twenty four hours after transfection, cells were treated with an inducer for 6 hours. Samples were collected and measured for SEAP activity. **Figure 4A** shows TNF-mediated transcriptional induction. Cells were treated with 100 ng/ml TNF; NF $\kappa$ B and NFAT signaling pathways exhibit 4-fold and 10-fold induction, respectively, when activated by TNF. **Figure 4B** shows serum-mediated induction of transcription factors in HEK 293 cells. Cells were fed with 10% serum. Multiple pathways were induced, when activated by serum. In both experiments, pTK-SEAP was used as the negative control. (RLU = relative light units)

15

## DETAILED DESCRIPTION OF THE INVENTION

The following terms have the definitions set below.

As used herein, "*cis*-element" means a genetic element having the ability to affect the activity of a gene on the same DNA molecule as that genetic element.

As used herein, "reporter" means a molecule (usually a protein) that is expressed in response to or as a result of a particular biological or molecular event.

As used herein, "cis-acting reporter" means a protein that is expressed in a measurable manner in response to a molecular event that takes place on the same DNA strand that the DNA which codes for the protein is located.



As used herein, "transcription factor" means any factor that controls genetic transcription.

As used herein, "signal transduction pathway" means the pathway or sequence of molecular events which converts an  
5 extracellular signal into cellular activity.

The present invention is directed to a series of *cis* element-reporter constructs and uses thereof.

In one embodiment of the present invention, there is provided a *cis* element-reporter construct comprising a *cis* element, a  
10 reporter gene and a promoter. Representative *cis* elements include nuclear factor- $\kappa$ B, heat responsive element, Myc, p53, nuclear factor of activated T cells, activator protein 1, serum responsive element, glucocorticoid responsive element and cAMP responsive element. In one preferred embodiment, the nuclear factor- $\kappa$ B has the sequence  
15 shown in SEQ ID No: 1. In one preferred embodiment, the heat responsive element has the sequence shown in SEQ ID No: 2. In one preferred embodiment, the Myc has the sequence shown in SEQ ID No: 3. In one preferred embodiment, the p53 has the sequence shown in SEQ ID No: 4. In one preferred embodiment, the nuclear factor of  
20 activated T cells has the sequence shown in SEQ ID No: 5. In one preferred embodiment, the activator protein 1 has the sequence shown in SEQ ID No: 6. In one preferred embodiment, the serum responsive element has the sequence shown in SEQ ID No: 7. In one preferred embodiment, the glucocorticoid responsive element has the sequence  
25 shown in SEQ ID No: 8. In one preferred embodiment, the cAMP responsive element has the sequence shown in SEQ ID No: 9 or SEQ ID No:10. Representative reporter genes include secreted alkaline phosphatase (SEAP), destabilized green fluorescent protein (d2EGFP) and luciferase (Luc). Representative promoters include thymidine

kinase promoter, gonadotropin  $\alpha$ -gene promoter, and other known by those having ordinary skill in this art.

In another embodiment of the present invention, there is provided a method of generating a *cis* element-reporter construct by linking a *cis* element to a promoter and then fusing the element to a reporter gene.

In yet another embodiment of the present invention, there is provided a method of monitoring activation of a transcription factor, comprising the steps of generating a vector comprising a *cis* element, a reporter gene and a promoter, wherein the *cis* element is a component of a DNA sequence to which said transcription factor binds to ensure activity; transfecting a cell line with the vector; and detecting expression of the reporter gene, wherein expression of the reporter gene indicates activation of the transcription factor. Representative cell lines include CHO, HEK 293, Saos2 and HeLa.

In still yet another embodiment of the present invention, there is provided a method of determining whether a gene is involved in a signal transduction pathway, comprising the steps of generating a vector comprising a *cis* element, a reporter gene and a promoter, wherein the *cis* element is a component of a DNA sequence which affects the expression of a test gene; expressing the vector in a cell line, wherein the cell line is associated with the signal transduction pathway; and detecting the expression of the reporter gene, wherein expression of the reporter gene indicates the involvement of the gene in the signal transduction pathway.

The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion.

**EXAMPLE 1**Cis-Element

The following *cis* elements were utilized for constructing  
5 cis-acting reporters:

NF- $\kappa$ B (nuclear factor- $\kappa$ B) having the sequence of GGGAA  
TTTCCGGGAATTTCCGGGAATTTCCGGGAATTTCC (SEQ ID No: 1);

HRE (heat responsive element) having the sequence of  
TAGAATGTTCTAGATCTAGAACATTCTAGCTAGAATGTTCTAG (SEQ ID No:  
10 2);

Myc having the sequence of CACGTGCACGTGCACGTGCACG  
TGCACGTGCACGTG (SEQ ID No: 3);

p53 having the sequence of TCGAGCTTGCCTGGACTTGCCTG  
GGCCAGATCTGTCGACGGAGGTCGAGCTTGCCTGGACTTGCCTGCCCACATCT  
15 GTCGACGGAGGTCGAGCTTGCCTGGACTTGCCTGCCCACATCTGTCGACGGAG  
GTCGAGCCTTGCCTGGACTTGCCTGCCA (SEQ ID No: 4);

NFAT (nuclear factor of activated T cells) having the  
sequence of GGAGGAAAACTGTTTCATACAGAAGGCGTGGAGGAAAAA  
CTGTTTCATACAGAAGGCGTGGAGGAAAACTGTTTCATACAGAAGGCGT  
20 (SEQ ID No: 5);

AP1 (activator protein 1) is an active protein and its activity  
is composed primarily by Jun/Fos heterodimers as well as Jun  
homodimers. The activity of AP1 responds to growth factors, such as  
phorbol ester (TPA), protein kinase C. Ap1 having the sequence of  
25 TGAGTCAGTGAGTCACTGACTCACTGACTCATGAGTCAGCTG  
ACTC (SEQ ID No: 6);

SRE (serum responsive element) having the sequence of  
GATGTCCATATTAGGACATCGATGTCCGAATATGGACATCGATGTCCATATT  
AGGACATC (SEQ ID No: 7);

GRE (glucocorticoid responsive element) having the sequence of GGTACATTTTGTCTAGAACAAAATGTACCGGTACATTTTGT TCT (SEQ ID No: 8);

CRE stands for cAMP responsive element. Two CRE elements were utilized: CRE3 (GCACCAGACAGTGACGTCAGCTGCCAGAT CCCATGGCCGTCATACTGTGACGTCTTTCAGACACCCCATTGACGTCAATGGG AGAAC: SEQ ID No: 9) and CRE5 (TCGAGCCCATGGCCGTCATACTGTGACG TCCCCATGGCCGTCATACTGTGACGTCCCCATGGCCGTCATACTGTGACGTCC CCATGGTCGTCATACTGTGACGTCCCCATGGCCGTCATACTGTGACGTC: SEQ ID No:10).

## **EXAMPLE 2**

### **Cis-Element/SEAP Constructs**

A number of *cis* elements were linked to a secreted alkaline phosphatase reporter wherein the secreted alkaline phosphatase was used as a reporter. One advantage of using SEAP as a reporter is that SEAP is secreted into the medium and the medium itself can be collected for the assay without disrupting the cells.

In a AP1-SEAP construct, the *cis* element in the construct contains six copies of AP1 binding sequence. In a SRE-SEAP construct, the construct contains three copies of SRE element. In a CRE-SEAP construct, the construct contains three different cAMP responsive elements, responding to cAMP, protein kinase A, forskolin etc. In a GRE-SEAP construct, the construct contains three copies of glucocorticoid responsive element and responds to the glucocorticoid-activated glucocorticoid receptor. In a HRE-SEAP construct, the construct contains three copies of heat responsive element. It responds to a number of types of physiological stress, such as heat. In an NF- $\kappa$ B-SEAP construct, the construct contains four copies of the NF-

κB binding sequence and responds to a number of stimuli, including TNF. In a NFAT-SEAP construct, the construct contains three copies of NFAT binding sequence and responds to PMA, ionomycin, PKC, etc. In a Myc-SEAP construct, the construct contains six copies of myc binding sequence and responds to myc activation. In a P53-SEAP construct, the construct contains two copies of p53 binding sequence, reporting p53 activation.

### **EXAMPLE 3**

10

#### **pNF-κB-SEAP Construct**

pNF-κB-SEAP (SEQ ID No:11) is especially designed for monitoring the activation of NF-κB signal transduction pathway (13-16). The restriction map of pNF-κB-SEAP is shown in Figure 1. Induction of the pathway enables endogenous NF-κB to bind to the kappa (κ) enhancer element (KB<sub>4</sub>) located in the promoter region of the vector, thus activating transcription of the reporter gene.

pNF-κB-SEAP allows expression of the secreted alkaline phosphatase reporter gene (17-19). The secreted alkaline phosphatase reporter coding sequence is followed by the SV40 late polyadenylation signal to ensure proper, efficient processing of the secreted alkaline phosphatase reporter transcript in eukaryotic cells. A synthetic transcription blocker (TB), composed of adjacent polyadenylation and transcription pause sites located upstream of the κ-enhancer, reduces background transcription (20). The vector backbone also contains an f1 origin for single-stranded DNA production, a pUC origin of replication, and an ampicillin resistance gene for propagation and selection in *E.coli*. The SEAP vectors incorporate a number of additional features that improve the sensitivity of secreted alkaline

phosphatase reporter by increasing the efficiency of secreted alkaline phosphatase reporter expression or that enhance the utility of the vectors. These include: the removal of the SV40 small-t intron, which can cause cryptic splicing and reduced expression in some genes and/or cell types (21-22); switching from the early to late polyadenylation signal of SV40, which typically causes a five-fold increase in mRNA levels (23); compact plasmid size; and removal of extraneous sequence from the 3' untranslated region of the SEAP mRNA.

10

#### **EXAMPLE 4**

##### **Cis-Element/EGFP Construct**

EGFP was also used as a reporter to analyze transcription factors. Destabilized EGFP (d2EGFP) is an autofluorescence protein wherein fluorescence emission of this protein does not require any additional substrates or cofactors. d2EGFP is an ideal reporter to be used in cell-based assays for high throughput drug screening as it allows real-time monitoring of activation of transcription factors. The detection can be made in multiple points.

A set of d2EGFP reporters with different cis-elements were generated, which are used for monitoring different transcription factors. Establishment of stable cell lines that express individual reporters expands application of this cis-element reporter in the cell-based high throughput drug screening.

An NF- $\kappa$ B-d2EGFP construct also is used to monitor the activation of NF- $\kappa$ B. This construct responds to a number of stimuli, including TNF and IL1. The AP1-d2EGFP construct is used to monitor the activation of AP1, which responds to growth factors. Additionally, an SRE-d2EGFP construct is used to monitor the activation of Elk1. This

30

construct responds to serum. The GRE-d2EGFP construct was used to monitor the activation of glucocorticoid receptors and responds to glucocorticoid activation. The CRE-d2EGFP construct was used to monitor CREB activation and responds to cAMP and cAMP-mediated protein kinase A (PKA). An NFAT-d2EGFP construct was used to monitor  $\text{Ca}^{++}$ -mediated NFAT activation which responds to  $\text{Ca}^{++}$ . The HSE-d2EGFP construct is used to monitor HSF activation and responds to a stress, such as heat. The P53-d2EGFP construct is used to monitor p53 activation, responding to DNA damage. The Myc-d2EGFP construct is used to monitor Myc activation, which responds to cell growth.

To construct the cis-acting d2EGFP reporters, the binding sequences of *cis* element was fused to the thymidine kinase (TK) promoter and d2EGFP gene. Resulting constructs are able to monitor the activation of a number of transcription factors. In CRE-d2EGFP transfected CHO cells, forskolin treatment elevates intracellular cAMP resulting in induction of d2EGFP. In NFAT-d2EGFP transfected HEK 293 cells, ionomycin and PMA treatment activates NFAT resulting in induction of d2EGFP. In GRE-d2EGFP transfected Saos2 cells, dexamethasone treatment activates glucocorticoid receptor (GR) resulting induction of d2EGFP.

## **EXAMPLE 5**

### **pNF- $\kappa$ B-d2EGFP Construct**

By linking NF- $\kappa$ B binding sequences and the thymidine kinase promoter, d2EGFP gene was able to detect TNF-mediated activation of NF- $\kappa$ B. pNF- $\kappa$ B-d2EGFP (SEQ ID No:12) was designed to monitor the activation of the NF- $\kappa$ B signal transduction pathway (13-16). The restriction map of pNF- $\kappa$ B-d2EGFP is shown in Figure 2.

Induction of the pathway enables endogenous NF- $\kappa$ B to bind to the kappa ( $\kappa$ ) enhancer element (KB<sub>4</sub>) located in the promoter region of the vector, thus activating transcription of the reporter gene.

d2EGFP is a destabilized variant of the enhanced green fluorescent protein. Unlike the original EGFP protein which is extremely stable, d2EGFP has a half-life of approximately two hours *in vivo* (24). Therefore, d2EGFP provides a more accurate measurement of transient activation by NF- $\kappa$ B than unmodified EGFP and is ideal for kinetic studies of gene activation. To construct d2EGFP, residues 422-461 of mouse ornithine decarboxylase (MODC) were fused to the C-terminus of EGFP. This region of the mouse ornithine decarboxylase contains a PEST amino acid sequence that targets the protein for degradation results in rapid protein turnover. pNF- $\kappa$ B-d2EGFP allows for real-time analysis of pNF- $\kappa$ B activation by either fluorescence microscopy or flow cytometry.

The d2EGFP coding sequence is followed by the SV40 late polyadenylation signal to ensure proper, efficient processing of the d2EGFP transcript in eukaryotic cells. A synthetic transcription blocker (TB), composed of adjacent polyadenylation and transcription pause sites located upstream of the  $\kappa$  enhancer reduces background transcription (20). The vector backbone also contains fl origin for single-stranded DNA production, a pUC origin of replication, and an ampicillin resistance gene for propagation and selection in *E.coli*. The d2EGFP vectors incorporate a number of features that improve the sensitivity of d2EGFP by increasing the efficiency of d2EGFP expression or that enhance the utility of the vectors. These include: the removal of the SV40 small-t intron, which can cause cryptic splicing and reduced expression in some genes and/or cell types (21-22); switching from the early to late polyadenylation signal of SV40, which typically



causes a five-fold increase in mRNA levels (23); compact plasmid size; and removal of extraneous sequence from the 3' untranslated region of the d2EGFP mRNA.

5

## **EXAMPLE 6**

### **CRE5-d2EGFP Construct**

pCRE5-d2EGFP (SEQ ID No:13) was constructed using the pTK-SEAP as the backbone. The thymidine kinase promoter was substituted by 5 copies of CREs (cAMP response elements) fused with gonadotropin  $\alpha$ -gene promoter (also called Thyroid-stimulating hormone TSH $\alpha$  subunit). The d2EGFP reporter gene was used instead of secreted alkaline phosphatase. This construct was designed to monitor the cAMP level changes *in vivo*. cAMP elevation within cells upon stimulation should result in the phosphorylation of CREB, which consequently induces the transcription and translation of the reporter gene d2EGFP. The destabilized form of EGFP eliminated the accumulation of EGFP in cells and better represented the changes of its expression level upon stimulation.

20

The stable cell line was generated by stably transfected pCRE5-d2EGFP together with pSV2-neo into CHO-K1 cells. Transfected cells were selected under G418 at a 500  $\mu$ g/ml concentration and individual clones were isolated and analyzed for forskolin induction. Three cell clones were selected which show a time course as well as the positive response to forskolin. Clone CRE5-dE #3 has the highest fluorescence intensity and is the easiest to detect induction, since it had five fold increase of fluorescence after induction. Clone CRE5-dE #38 has the highest fold of induction (10 fold) and the lowest background EGFP expression. However, after induction, Clone CRE5-dE

25

#38 had a lower fluorescent intensity than that of clone #3. Clone #38-2 is an intermediate of above two clones, having an 8 fold fluorescence induction and, after induction, the fluorescence intensity was at the similar level of clone #3. These cell clones can be used in cell based high-throughput screening in search of factors involved in cAMP signal transduction pathway.

### **EXAMPLE 7**

#### **Cis-Element/Luciferase Construct**

Luciferase was also used as a reporter to analyze transcription factors. Luciferase is one of most sensitive reporters and is widely used in the art. Due to its sensitivity and popularity, luciferase was selected to link to a *cis* element to form a *cis*-element/luciferase construct for monitoring the activation of transcription factors.

pNF- $\kappa$ B-Luciferase (pNF- $\kappa$ B-Luc, SEQ ID No:14) contains a firefly luciferase gene and is designed for monitoring the activation of the NF- $\kappa$ B signal transduction pathway (13-16). The restriction map of pNF- $\kappa$ B-Luc is shown in Figure 3. Induction of the pathway enables endogenous NF- $\kappa$ B to bind to the kappa ( $\kappa$ ) enhancer element (KB<sub>4</sub>) located in the promoter region of the vector, thus activating transcription of the reporter gene. The luciferase coding sequence is followed by the SV40 late polyadenylation signal to ensure proper, efficient processing of the luciferase transcript in eukaryotic cells. A synthetic transcription blocker, composed of adjacent polyadenylation and transcription pause sites located upstream of the  $\kappa$ -enhancer, reduces background transcription (20). The vector backbone also contains an f1 origin for single-stranded DNA production, a pUC origin

of replication, and an ampicillin resistance gene for propagation and selection in *E.coli*.

The pNF- $\kappa$ B vectors incorporate a number of features that improve the sensitivity of luciferase by increasing the efficiency of luciferase expression or that enhance the utility of the vectors. These include: the removal of the SV40 small-t intron, which can cause cryptic splicing and reduced expression in some genes and/or cell types (21-22); switching from the early to late polyadenylation signal of SV40, which typically causes a five-fold increase in mRNA levels (23); compact plasmid size; and removal of extraneous sequence from the 3' untranslated region of the luciferase mRNA.

### **EXAMPLE 8**

#### **Application of Cis-Acting Reporters in Assaying the Activity of Transcription Factors in Signal Pathways**

The activity assay of the present invention may be carried out in a 12 or 24 well plate. This assay can be applied to a lot of studies, such as examining whether a new target gene links to any signal pathways. The assay can be used for establishing a functional status profile of transcription factors for evaluating any biological difference *in vivo*, such as normal cells *versus* differentiated, apoptotic, and cancer cells. The assays of the present invention can also be used to examine and quickly identify biological markers for human diseases. Drug validation is another application of the present assay.

Figure 4 demonstrates mediated induction of transcription factors in HEK 293 cells. Figure 4A shows tumor necrosis factor (TNF)-mediated transcriptional induction. Human embryonic kidney (HEK 293) cells were transfected individually with each reporter vector and

incubated for 24 hr in serum-free media. After incubation with TNF for 6 hr, samples were collected and measured for SEAP activity. The transcriptional profile reveals that, as expected, TNF activated the NF $\kappa$ B and the NFAT signaling pathways, which are known to play pivotal roles in the transcription of genes involved in the immune response (1-4). Other transcription factors and controls were unaffected by TNF stimulation.

Figure 4B shows serum-mediated induction of transcription factors HEK293 cells. After transfection, cells were cultured in serum-free media for 24 hr and incubated with 10% serum for 6 hr. Samples were collected and measured for SEAP activity as described above. Unlike the TNF-mediated activation of specific pathways (Figure 4A), serum activates multiple transcription factors that induce a variety of signaling events after stimulation.

The following references were cited herein.

1. Chalfie, et al., (1994) Science 263: 802-805.
2. Prasher, et al., (1992) Gene 111: 229-233.
3. Inouye, et al., (1994) FEBS Letters 341: 277-280.
4. Wang, et al., (1994) Nature 369: 400-403.
5. Yang, et al., (1996) Nucleic Acid Res. 24(22): 4592-4593.
6. Cormack, et al., (1996) Gene 173: 33-38.
7. Hass, et al., (1996) Curr. Biol. 6: 315-324.
8. Galbraith, et al., (1995) Methods. Cell Biol. 50: 1-12.
9. Living Colors Destabilized EGFP Vectors (April 1998), CLONTECHniques XIII(2): 16-17.
10. Berger, et al., (1998) Gene 66: 1-10.
11. Kain, et al., (1997) Methods Mol. Biol. 63: 49-60.
12. Cullen, et al., (1992) Methods Enzymol. 216: 362-368.
13. Baeuerle, P.A. & Baltimore, D. (1996) Cell 87:13-20
14. Baeuerle, P.A. (1998) Curr. Biol. 8:R19-R2

15. Peltz, G. (1997) Curr. Opin. Biotechnol. 8:467-473
16. Israel, A. (1997) Nature 388: 519-521
17. Berger, J. et al. (1998) Gene 66:1-10
18. Kain, S.R. (1997) Methods Mol. Biol. 63:49-60
- 5 19. Cullen, B.R. & Malim, M.H. (1992) Methods Enzymol 216:362-368
20. Eggermont, J. & Proudfoot, N. (1993) EMBO J. 12:2539-2548
21. Huang, M.T.F. & Gorman, C.M. (1990) Mol. Cell. Biol. 10:1805-1810
- 10 22. Evans, M.J. & Scarpulla, R.C. (1989) Gene 84:135-142
23. Carswell, S. & Alwine, J.C. (1989) Mol. Cell. Biol. 9:5248-4258

Any patents or publications mentioned in this specification are indicative of the levels of those skilled in the art to which the invention pertains. These patents and publications are herein  
15 incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

One skilled in the art will readily appreciate that the present invention is well adapted to carry out the objects and obtain  
20 the ends and advantages mentioned, as well as those inherent therein. The present examples along with the methods, procedures, treatments, molecules, and specific compounds described herein are presently representative of preferred embodiments, are exemplary, and are not  
25 intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention as defined by the scope of the claims.

**WHAT IS CLAIMED IS:**

1. A construct for measuring transcription, comprising:  
a *cis* element, wherein said *cis* element is selected from the  
5 group consisting of nuclear factor-kB having the sequence shown in  
SEQ ID No: 1, heat responsive element having the sequence shown in  
SEQ ID No: 2, Myc having the sequence shown in SEQ ID No: 3, p53  
having the sequence shown in SEQ ID No: 4, nuclear factor of activated  
T cells having the sequence shown in SEQ ID No: 5, alkaline  
10 phosphatase 1 having the sequence shown in SEQ ID No: 6, serum  
responsive element having the sequence shown in SEQ ID No: 7,  
glucocorticoid responsive element having the sequence shown in SEQ  
ID No: 8 and cAMP responsive element having the sequence selected  
from the group consisting of SEQ ID No: 9 and SEQ ID No: 10;  
15 a reporter gene; and  
a promoter.

2. The construct of claim 1, wherein said reporter gene  
20 is selected from the group consisting of secreted alkaline phosphatase,  
destabilized green fluorescent protein and luciferase.

3. The construct of claim 1, wherein said promoter is  
25 selected from the group consisting of thymidine kinase promoter and  
gonadotropin  $\alpha$ -gene promoter.

4. A vector comprising the construct of claim 1.  
30

5. The vector of claim 4, wherein said vector is pNF- $\kappa$ B-SEAP.

5 6. The vector of claim 4, wherein said vector is pNF- $\kappa$ B-d2EGFP.

7. The vector of claim 4, wherein said vector is pNF- $\kappa$ B-Luc.  
10

8. A method of monitoring activation of a transcription factor, comprising the steps of:

15 generating the construct of claim 1, wherein said *cis*-element is a transcription factor;

transforming a cell line with a vector containing said construct; and

20 detecting the expression of said reporter gene, wherein expression of said reporter gene indicates the activation of said transcription factor.

9. A method of monitoring activation of a transcription factor, comprising the steps of:

25 generating the construct of claim 2;

transforming a cell line with a vector containing said construct; and

detecting the expression of said reporter gene, wherein expression of said reporter gene indicates the activation of said transcription factor.

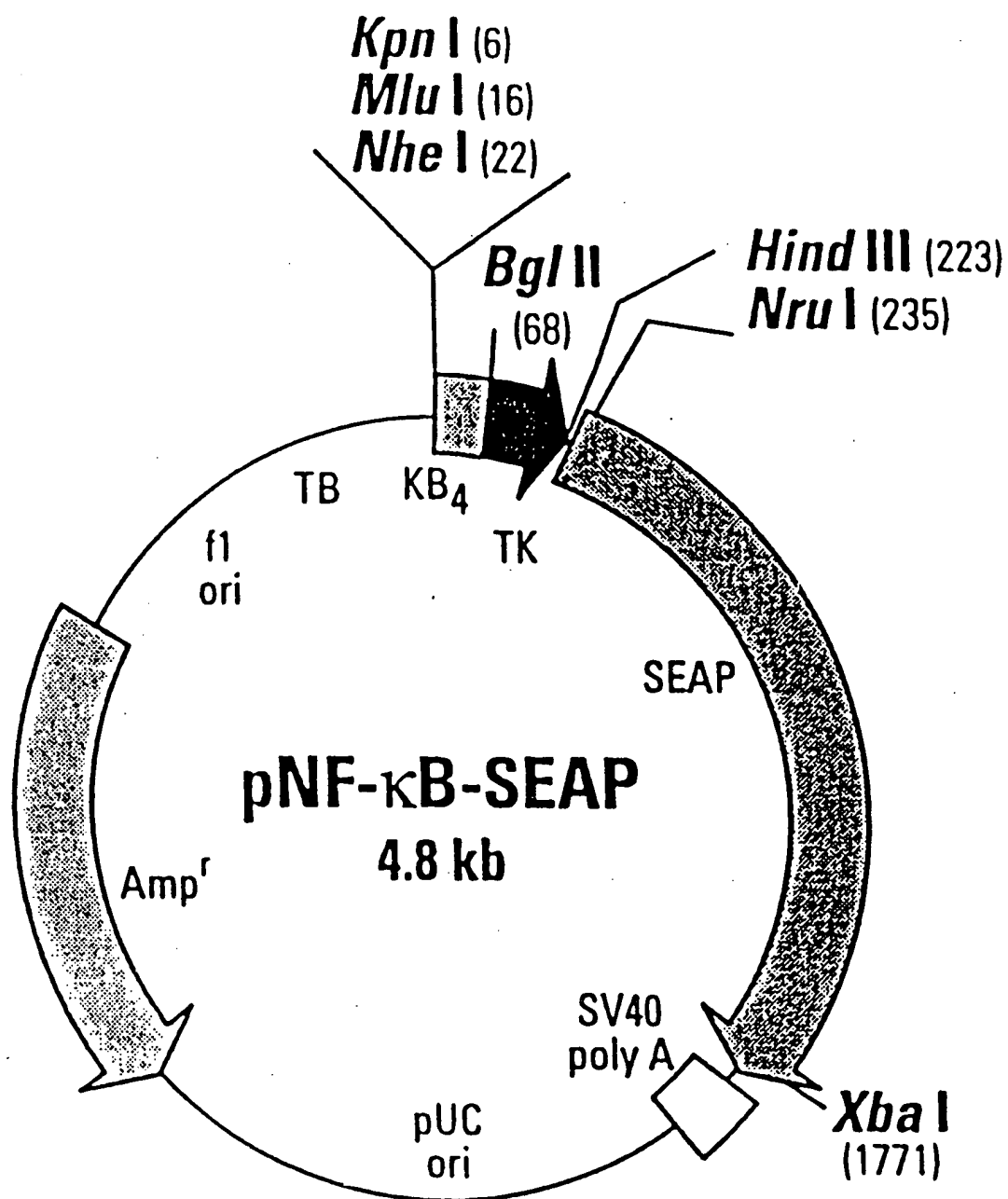
5

10. The method of claim 9, wherein said cell line is selected from the group consisting of CHO, HEK 293 and Saos2.

10

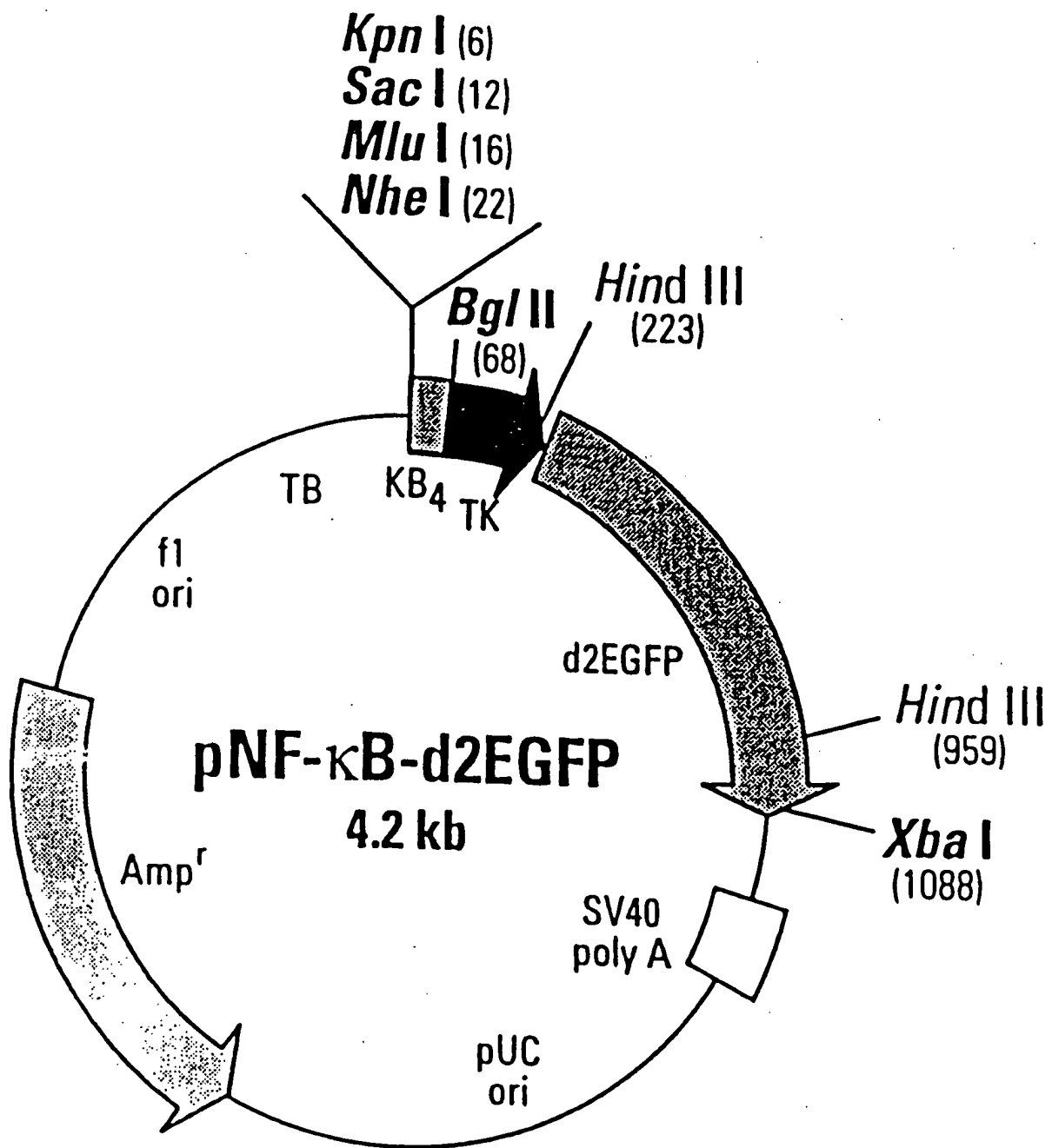
11. The method of claim 9, wherein said construct is selected from the group consisting of pNF- $\kappa$ B-SEAP, pNF- $\kappa$ B-d2EGFP and pNF- $\kappa$ B-Luc.





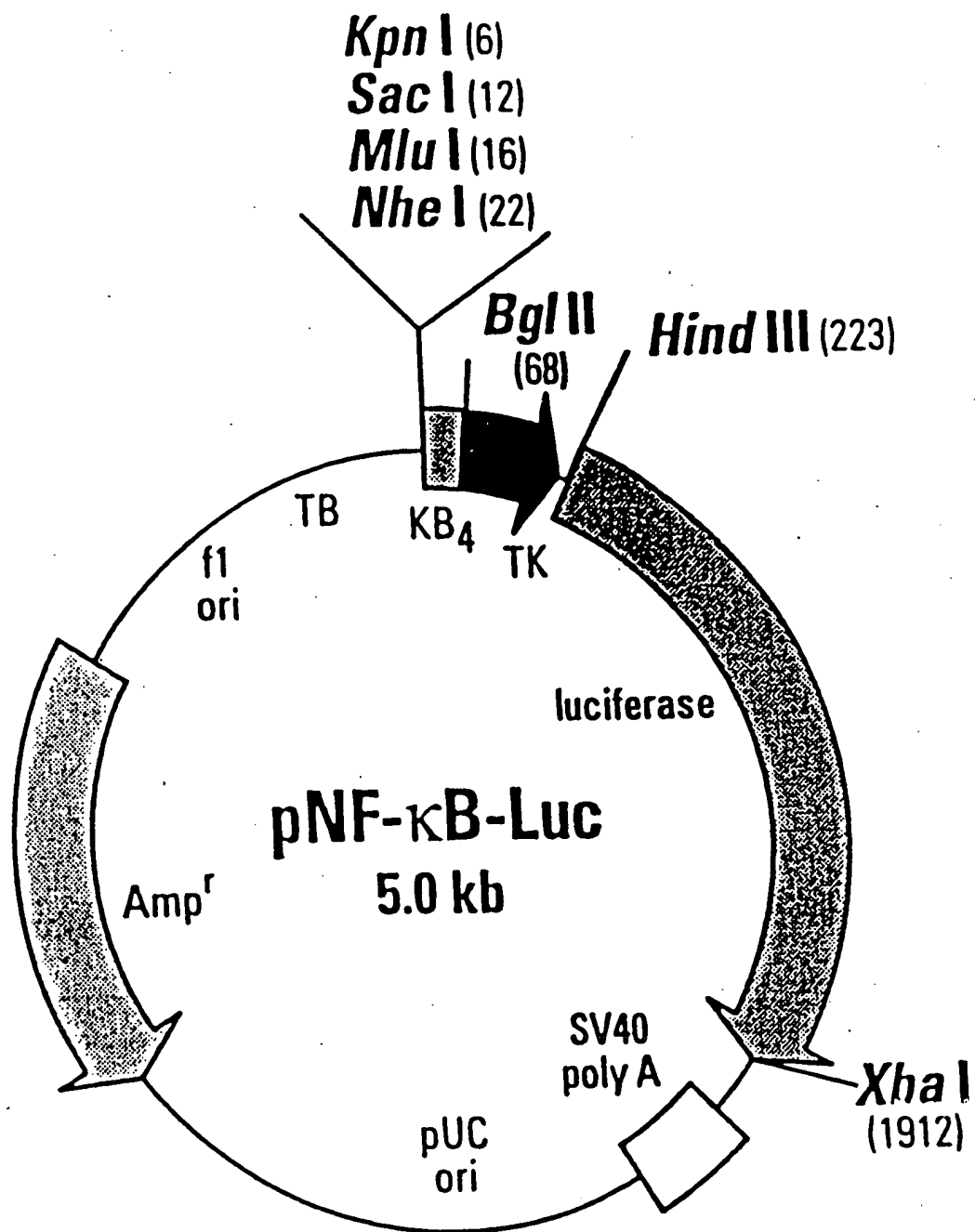
TB = Transcription Blocker

FIG. 1



TB = Transcription Blocker

FIG. 2



TB = Transcription Blocker

FIG. 3

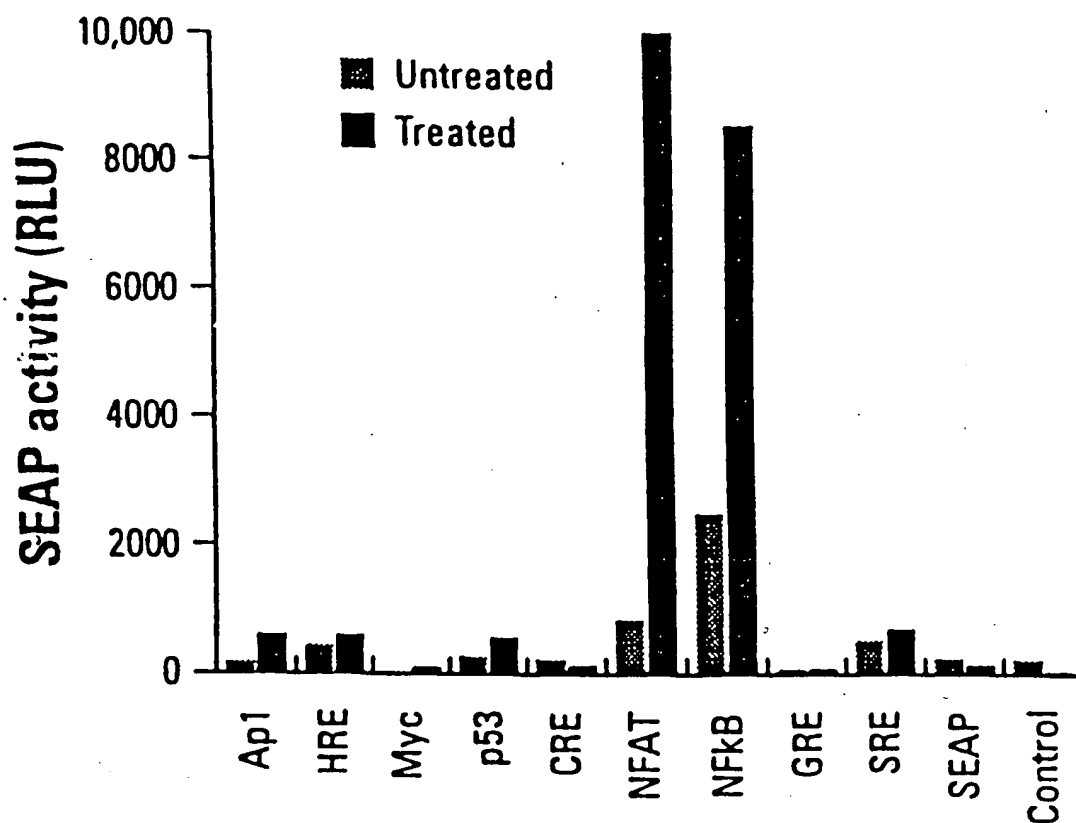


FIG. 4A

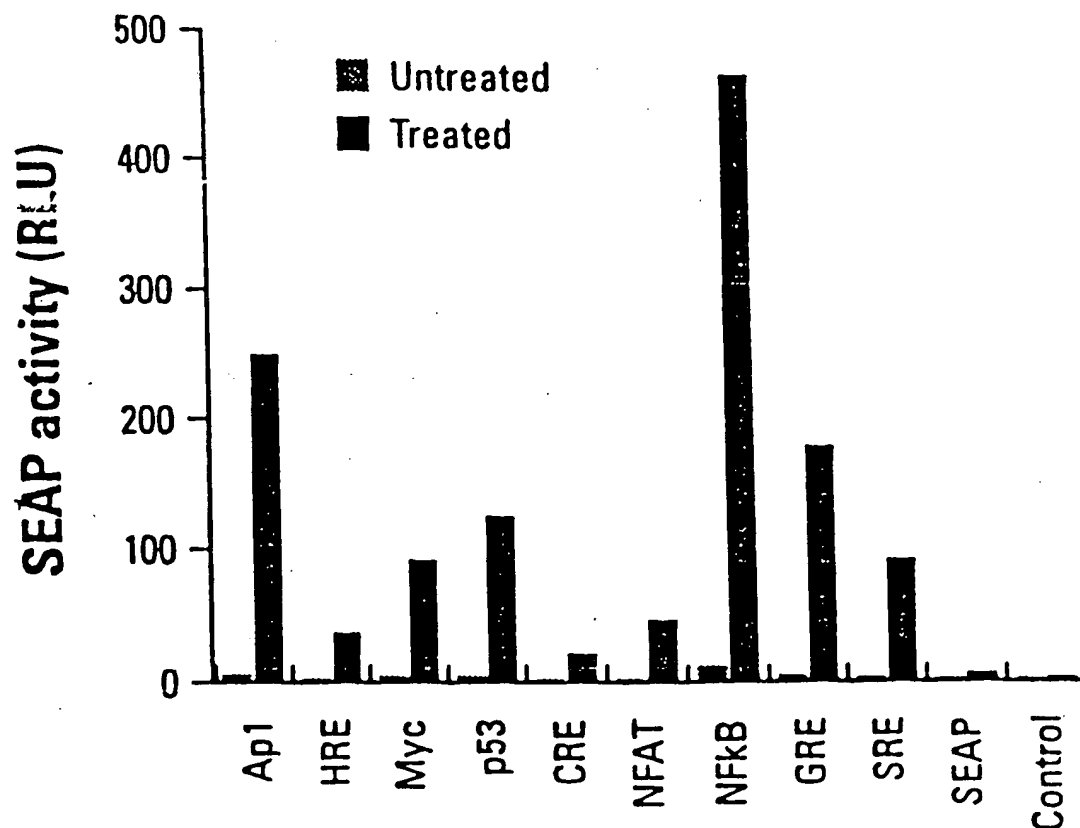


FIG. 4B

## SEQUENCE LISTING

<110> Li, Xianqiang  
Zhao, Xiaoning  
Fang, Yu  
Jiang, Xin  
Duong, Tommy  
Kain, Steve R.

<120> Cis-Element Reporter Constructs and Uses Thereof

<130> D6170PCT

<141> 1999-12-01

<150> US 09/206,887

<151> 1998-12-08

<160> 14

<210> 1

<211> 40

<212> DNA

<213> artificial sequence

<220>

<223> NF-κB element

<400> 1

gggaatttcc gggaatttcc gggaatttcc gggaatttcc 40

<210> 2

<211> 43

<212> DNA

<213> artificial sequence

<220>

<223> heat responsive element (HRE)

<400> 2

tagaatgttc tagatctaga acattctagc tagaatgttc tag 43

<210> 3

<211> 36

<212> DNA

<213> artificial sequence

<220>

<223> Myc element

<400> 3  
 cacgtgcacg tgcacgtgca cgtgcacgtg cacgtg 36

<210> 4  
 <211> 157  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <223> p53 cis element  
 <400> 4

tcgagcttgc ctggacttgc ctgggccaga tctgtcgacg gaggtcgagc 50  
 ttgcctggac ttgcctgccc acatctgtcg acggaggtcg agcttgcttg 100  
 gacttgcttg cccacatctg tcgacggagg tcgagccttg cctggacttg 150  
 cctgcca 157

<210> 5  
 <211> 90  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <223> nuclear factor of activated T cells (NF-AT)  
 <400> 5

ggaggaaaaa ctgtttcata cagaaggcgt ggaggaaaaa ctgtttcata 50  
 cagaaggcgt ggaggaaaaa ctgtttcata cagaaggcgt 90

<210> 6  
 <211> 46  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <223> alkaline phosphatase 1 (AP1)  
 <400> 6

tgagtcagtg agtcactgac tcactgactc atgagtcagc tgactc 46

<210> 7  
 <211> 60  
 <212> DNA

<213> artificial sequence

<220>

<223> serum responsive element (SRE)

<400> 7

gatgtccata ttaggacatc gatgtccgaa tatggacatc gatgtccata 50  
ttaggacatc 60

<210> 8

<211> 45

<212> DNA

<213> artificial sequence

<220>

<223> glucocorticoid responsive element (GRE)

<400> 8

ggtacatttt gttctagaac aaaatgtacc ggtacatttt gttct 45

<210> 9

<211> 87

<212> DNA

<213> artificial sequence

<220>

<223> cAMP responsive element (CRE3)

<400> 9

gcaccagaca gtgacgtcag ctgccagatc ccatggccgt cataactgtga 50  
cgtctttcag acacccatt gacgtcaatg ggagaac 87

<210> 10

<211> 130

<212> DNA

<213> artificial sequence

<220>

<223> cAMP responsive element (CRE5)

<400> 10

tcgagcccat ggccgtcata ctgtgacgtc cccatggccg tcataactgtg 50  
acgtcccat ggccgtcata ctgtgacgtc cccatggctg tcataactgtg 100  
acgtcccat ggccgtcata ctgtgacgtc 130

<210> 11  
 <211> 4846  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <223> synthesized pNF-KB-SEAP construct  
 <400> 11

```

ggtaccgagc tcttacgcgt gctagcggga atttccggga atttccggga 50
atttccggga atttccagat ctgccgcccc gactgcatct gcgtgttcga 100
attcgccaat gacaagacgc tgggcgggggt ttgtgtcatc atagaactaa 150
agacatgcaa atatatcttct tccgggggaca ccgccagcaa acgcgagcaa 200
cggggccacgg ggatgaagca gaagcttcga atcgcggaatt cgcccaccat 250
gctgctgctg ctgctgctgc tgggcctgag gctacagctc tccctgggca 300
tcatcccagt tgaggaggag aacccggact tctggaaccg cgaggcagcc 350
gaggccctgg gtgccgcaa gaagctgcag cctgcacaga cagccgcaa 400
gaacctcatc atcttctctg gcgatgggat gggggtgtct acggtgacag 450
ctgccaggat cctaaaaggg cagaagaagg acaaactggg gcctgagata 500
cccctggcca tggaccgctt cccatatgtg gctctgtcca agacatacaa 550
tgtagacaaa catgtgccag acagtggagc cacagccacg gcctacctgt 600
gcggggtcaa gggcaacttc cagaccattg gcttgagtgc agccgcccgc 650
ttaaccagt gcaacacgac acgcggaac gaggtcatct ccgtgatgaa 700
tcgggccaag aaagcaggga agtcagtggg agtggttaacc accacacgag 750
tgcagcacgc ctgccagcc ggcacctacg cccacacggt gaaccgcaac 800
tggtactcgg acgccgacgt gcctgcctcg gcccgccagg aggggtgcca 850
ggacatcgct acgcagctca tctccaacat ggacattgac gtgatacctag 900
gtggaggccg aaagtacatg tttcgcatgg gaaccccaga ccctgagtac 950
ccagatgact acagccaagg tgggaccagg ctggacggga agaactctggt 1000
gcaggaatgg ctggcgaagc gccagggtgc ccggtatgtg tggaaccgca 1050
ctgagctcat gcaggcttcc ctggaccctg ctgtgaccca tctcatgggt 1100
ctctttgagc ctggagacat gaaatacgag atccaccgag actccacact 1150
ggaccctctc ctgatggaga tgacagaggc tgccctgcgc ctgctgagca 1200
ggaacccccg cggcttcttc ctcttcgtgg aggggtggtcg catcgaccat 1250
ggtcatcatg aaagcagggc ttaccgggca ctgactgaga cgatcatgtt 1300
cgacgacgcc attgagaggg cgggccagct caccagcgag gaggacacgc 1350
tgagcctcgt cactgccgac cactcccacg tcttctcctt cggaggctac 1400
cccctgcgag ggagctccat cttcgggctg gccctggca agggccggga 1450
caggaaggcc tacacgggtc tcctatacgg aaacgggtcca ggctatgtgc 1500
tcaaggacgg cgcccggccg gatgttaccg agagcgagag cgggagcccc 1550
  
```



gagtatcggc agcagtcagc agtgcccctg gacgaagaga cccacgcagg 1600  
 cgaggacgtg gcggtgttcg cgcgcggccc gcaggcgcac ctggttcacg 1650  
 gcgtgcagga gcagaccttc atagcgcacg tcatggcctt cgccgcctgc 1700  
 ctggagccct acaccgctg cgacctggcg cccccgcgg gcaccaccga 1750  
 cgccgcgcac ccgggttact ctagagtcgg ggccggccggc cgcttcgagc 1800  
 agacatgata agatacattg atgagtttgg acaaaccaca actagaatgc 1850  
 agtgaaaaaa atgctttatt tgtgaaattt gtgatgctat tgctttattt 1900  
 gtaaccatta taagctgcaa taaacaagtt aacaacaaca attgcattca 1950  
 ttttatgttt caggttcagg gggaggtgtg ggaggttttt taaagcaagt 2000  
 aaaacctcta caaatgtggg aaaatcgata aggatccgtc gaccgatgcc 2050  
 cttgagagcc ttcaaccagc tcagctcctt ccggtggggc cggggcatga 2100  
 ctatcgtcgc cgcacttatg actgtcttct ttatcatgca actcgtagga 2150  
 caggtgcggc cagcgtctct ccgcttcctc gctcactgac tcgctgcgct 2200  
 cggctcgttcg gctgcggcga gcggtatcag ctactcaaa ggcggtaata 2250  
 cggttatcca cagaatcagg ggataacgca ggaaagaaca tgtgagcaaa 2300  
 aggccagcaa aaggccagga accgtaaaaa ggccgcgttg ctggcgtttt 2350  
 tccataggct ccgccccctt gacgagcatc aaaaaaatcg acgctcaagt 2400  
 cagaggtggc gaaacccgac aggactataa agataccagg cgtttccccc 2450  
 tggaagctcc ctcgctgcgct ctctgttcc gacctgccg cttaccggat 2500  
 acctgtccgc ctttctccct tcgggaagcg tggcgctttc tcatagctca 2550  
 cgctgtaggt atctcagttc ggtgtaggtc gttcgctcca agctgggctg 2600  
 tgtgcacgaa cccccgctt agcccgaccg ctgcgcctta tccggtaact 2650  
 atcgtcttga gtccaacccg gtaagacacg acttatcgcc actggcagca 2700  
 gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 2750  
 gttcttgaag tgggtggccta actacggcta cactagaagg acagtatttg 2800  
 gtatctgcgc tctgctgaag ccagttacct tcggaaaaag agttggtagc 2850  
 tcttgatccg gcaaacaac caccgctggg agcggtggtt tttttgtttg 2900  
 caagcagcag attacgcgca gaaaaaaagg atctcaagaa gatcctttga 2950  
 tcttttctac ggggtctgac gctcagtgga acgaaaactc acgttaaggg 3000  
 attttggtca tgagattatc aaaaaggatc ttcacctaga tccttttaaa 3050  
 ttaaaaatga agtttttaaat caatctaaag tatatatgag taaacttggg 3100  
 ctgacagtta ccaatgctta atcagtgagg cacctatctc agcgatctgt 3150  
 ctatttcgtt catccatagt tgctgactc cccgtcgtgt agataactac 3200  
 gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 3250  
 acccagctc accggtcca gatttatcag caataaacca gccagccgga 3300  
 agggccgagc gcagaagtgg tcctgcaact ttatccgctt ccatccagtc 3350  
 tattaattgt tgccgggaag ctagagtaag tagttcgcca gttaatagtt 3400  
 tgcgcaacgt tgttgccatt gctacaggca tcgtgggtgtc acgctcgtcg 3450

```

tttggtatgg cttcattcag ctccggttcc caacgatcaa ggcgagttac 3500
atgatcccc atgttgtgca aaaaagcggg tagctccttc ggtcctccga 3550
tcgttgtcag aagtaagttg gccgcagtgt tatcactcat ggttatggca 3600
gcactgcata attctcttac tgtcatgcc tccgtaagat gcttttctgt 3650
gactggtgag tactcaacca agtcattctg agaatagtgt atgcggcgac 3700
cgagttgctc ttgcccggcg tcaatacggg ataataccgc gccacatagc 3750
agaactttaa aagtgctcat cattggaaaa cgttcttcgg ggcgaaaact 3800
ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg 3850
cacccaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga 3900
gcaaaaacag gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg 3950
gaaatgttga atactcatac tcttcctttt tcaatattat tgaagcattt 4000
atcagggtta ttgtctcatg agcggataca tatttgaatg tatttagaaa 4050
aataaacaaa taggggttcc gcgcacattt ccccgaaaag tgccacctga 4100
cgcgccctgt agcggcgcat taagcgcggc ggggtgtggtg gttacgcgca 4150
gcgtgaccgc tacacttgcc agcgccctag cgcccgtcc tttcgctttc 4200
ttcccttctt ttctcgccac gtctgcggc tttccccgtc aagctctaaa 4250
tcgggggctc cttttagggg tccgatttag tgctttacgg cacctcgacc 4300
ccaaaaaact tgattagggt gatggttcac gtagtgggcc atcgccctga 4350
tagacggttt ttcgcccttt gacgttgagg tccacgttct ttaatagtgg 4400
actcttggtc caaactggaa caaactcaa ccctatctcg gtctattctt 4450
ttgatttata agggattttg ccgatttcgg cctattgggt aaaaaatgag 4500
ctgatttaac aaaaatttaa cgcgaaattt aacaaaatat taacgtttac 4550
aatttcccat tcgccattca ggctgcgcaa ctggtgggaa gggcgatcgg 4600
tgcgggcctc ttcgctatta cgccagccca agctaccatg ataagtaagt 4650
aatattaagg tacgggaggt acttgagcg gccgcaataa aatatcttta 4700
ttttcattac atctgtgtgt tggttttttg tgtgaatcga tagtactaac 4750
atacgtctc catcaaaaca aaacgaaaca aaacaaacta gcaaaatagg 4800
ctgtccccag tgcaagtgca ggtgccagaa catttctcta tcgata 4846

```

<210> 12

<211> 4163

<212> DNA

<213> artificial sequence

<220>

<223> synthesized pNF-KB-d2EGFP construct

<400> 12

```

ggtagccgagc tcttacgcgt gctagcggga atttccggga atttccggga 50

```

atttccggga	atttccagat	ctgcgcggcc	gactgcatct	gcgtgttcga	100
attcgccaat	gacaagacgc	tgggcgggggt	ttgtgtcatc	atagaactaa	150
agacatgcaa	atatatttct	tccgggggaca	ccgccagcaa	acgcgagcaa	200
cggggccacgg	ggatgaagca	gaagcttccc	cgaattccgg	atggtgagca	250
agggcgagga	gctgttcacc	ggggtggtgc	ccatcctggt	cgagctggac	300
ggcgacgtaa	acggccacaa	gttcagcgtg	tccggcgagg	gcgagggcga	350
tgccacctac	ggcaagctga	ccctgaagtt	catctgcacc	accggcaagc	400
tgcccgtgcc	ctggcccacc	ctcgtgacca	ccctgacctt	cggcgtgcag	450
tgcttcagcc	gctaccccga	ccacatgaag	cagcacgact	tcttcaagtc	500
cgccatgccc	gaaggctacg	tccaggagcg	caccatcttc	ttcaaggacg	550
acggcaacta	caagaccgcg	gccgaggtga	agttcgaggg	cgacaccctg	600
gtgaaccgca	tcgagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	650
cctggggcac	aagctggagt	acaactacaa	cagccacaac	gtctatatca	700
tggccgacaa	gcagaagaac	ggcatcaagg	tgaacttcaa	gatccgccac	750
aacatcgagg	acggcagcgt	gcagctcgcc	gaccactacc	agcagaacac	800
ccccatcggc	gacggccccg	tgctgctgcc	cgacaaccac	tacctgagca	850
cccagtcggc	cctgagcaaa	gaccccaacg	agaagcgcg	tcacatggtc	900
ctgctggagt	tcgtgaccgc	cgccgggata	actctcggca	tggacgagct	950
gtacaagaag	cttagccatg	gcttcccggc	ggaggtggag	gagcaggatg	1000
atggcacgct	gcccattgtct	tgtgcccagg	agagcgggat	ggaccgtcac	1050
cctgcagcct	gtgcttctgc	taggatcaat	gtgtagtcta	gagtcggggc	1100
ggccggccgc	ttcgagcaga	catgataaga	tacattgatg	agtttggaca	1150
aaccacaact	agaatgcagt	gaaaaaaatg	ctttatttgt	gaaatttgtg	1200
atgctattgc	tttatttgta	accattataa	gctgcaataa	acaagttaac	1250
aacaacaatt	gcattcattt	tatgtttcag	gttcaggggg	aggtgtggga	1300
ggttttttta	agcaagtaaa	acctctacaa	atgtggtaaa	atcgataagg	1350
atccgtcgac	cgatgccctt	gagagccttc	aaccagtcga	gctccttccg	1400
gtgggcgcgg	ggcatgacta	tcgtcgccgc	acttatgact	gtcttcttta	1450
tcatgcaact	cgtaggacag	gtgccggcag	cgctcttccg	cttcctcgct	1500
cactgactcg	ctgcgctcgg	tcgttcgggt	gcggcgagcg	gtatcagctc	1550
actcaaaggc	ggtaatacgg	ttatccacag	aatcagggga	taacgcagga	1600
aagaacatgt	gagcaaaagg	ccagcaaaag	gccaggaacc	gtaaaaaggc	1650
cgcgttgctg	gcgtttttcc	ataggctccg	ccccctgac	gagcatcaca	1700
aaaatcgacg	ctcaagtcag	aggtggcgaa	acccgacagg	actataaaga	1750
taccaggcgt	ttccccctgg	aagctccctc	gtgcgctctc	ctgttccgac	1800
cctgccgctt	accggatacc	tgtccgcctt	tctcccttcg	ggaagcgtgg	1850
cgctttctca	tagctcacgc	tgtaggtatc	tcagttcggt	gtaggtcggt	1900
cgctccaagc	tgggctgtgt	gcacgaacct	ccggttcagc	ccgaccgctg	1950

cgccttatcc ggtaactatc gtcttgagtc caacccggta agacacgact 2000  
tategccact ggcagcagcc actggtaaca ggattagcag agcgaggat 2050  
gtaggcggtg ctacagagtt cttgaagtgg tggcctaact acggctacac 2100  
tagaaggaca gtatttggtg tctgcgctct gctgaagcca gttaccttcg 2150  
gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc 2200  
ggtggttttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc 2250  
tcaagaagat cctttgatct tttctacggg gtctgacgct cagtggaaacg 2300  
aaaactcacg ttaagggatt ttgggtcatga gattatcaaa aaggatcttc 2350  
acctagatcc ttttaaatta aaaatgaagt tttaaataca tctaaagtat 2400  
atatgagtaa acttggtctg acagttacca atgcttaatc agtgaggcac 2450  
ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc 2500  
gtcgtgtaga taactacgat acgggagggc ttaccatctg gccccagtgc 2550  
tgcaatgata ccgcgagacc cacgctcacc ggctccagat ttatcagcaa 2600  
taaaccagcc agccggaagg gccgagcgca gaagtgggtcc tgcaacttta 2650  
tccgcctcca tccagtctat taattggtgc cgggaagcta gagtaagtag 2700  
ttcgccagtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg 2750  
tgggtgtcacg ctcgctggtt ggtatggctt cattcagctc cggttcccaa 2800  
cgatcaaggc gagttacatg atcccccatg ttgtgcaaaa aagcggttag 2850  
ctccttcggt cctccgatcg ttgtcagaag taagttggcc gcagtgttat 2900  
cactcatggt tatggcagca ctgcataatt ctcttactgt catgccatcc 2950  
gtaagatgct tttctgtgac tgggtgagta tcaaccaagt cattctgaga 3000  
atagtgtatg cggcgaccga gttgctcttg cccggcgta atacgggata 3050  
ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt 3100  
tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc 3150  
gatgtaaccc actcgtgcac ccaactgatc ttcagcatct tttactttca 3200  
ccagcgtttc tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag 3250  
ggaataaggg cgacacggaa atgttgaata ctcatactct tcctttttca 3300  
atattattga agcatttatc aggggttattg tctcatgagc ggatacatat 3350  
ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc 3400  
cgaaaagtgc cacctgacgc gccctgtagc ggcgcatata gcgcggcggg 3450  
tgtggtggtt acgcgcagcg tgaccgctac acttgccagc gccctagcgc 3500  
ccgctccttt cgctttcttc ccttccttcc tcgccacggt cgccggcctt 3550  
ccccgtcaag ctctaaatcg ggggctccct ttagggttcc gatttagtgc 3600  
tttacggcac ctcgacccca aaaaacttga ttaggggtgat ggttcacgta 3650  
gtggggccatc gccctgatag acggtttttc gccctttgac gttggagtcc 3700  
acgttcttta atagtggact cttgttccaa actggaacaa cactcaaccc 3750  
tatctcggtc tattcttttg atttataagg gattttgccg atttcggcct 3800  
attggttaaa aatgagctg atttaacaaa aatttaacgc gaattttaac 3850

```

aaaatattaa cgtttacaat ttcccattcg ccattcaggc tgcgcaactg 3900
ttgggaaggg cgatecgggtgc gggcctcttc gctattacgc cagcccaagc 3950
taccatgata agtaagtaat attaaggtac gggaggtact tggagcggcc 4000
gcaataaaat atctttatctt tcattacatc tgtgtgttgg ttttttgtgt 4050
gaatcgatag tactaacata cgctctccat caaaacaaaa cgaaacaaaa 4100
caaactagca aaataggctg tccccagtg aagtgacagg gccagaacat 4150
ttctctatcg ata 4163

```

```

<210>      13
<211>      4300
<212>      DNA
<213>      artificial sequence
<220>
<223>      synthesized pCRE5-d2EGFP
<400>      13

```

```

ggtaccgagc tcttacgcgt gctagcggga atttccggga atttccggga 50
atttccggga atttccagat cctcgagccc atggccgtca tactgtgacg 100
tccccatggc cgtcatactg tgacgtcccc atggccgtca tactgtgacg 150
tccccatggt cgtcatactg tgacgtcccc atggccgtca tactgtgacg 200
tcccgcggtc gactaccctt caatcattgg atggaatttc ctgttgatcc 250
cagggccttag atgcaggtgg aaacactctg ctggtataaa agcaggtgag 300
gacttcatta actgcagtta ctgagaactc ataagacgaa gctaaaatcc 350
ctcttcggaa gcttccccga attccggatg gtgagcaagg gcgaggagct 400
gttcaccggg gtggtgcccc tcctgggtcga gctggacggc gacgtaaacg 450
gccacaagtt cagcgtgtcc ggcgaggggc agggcgatgc cacctacggc 500
aagctgaccc tgaagttcat ctgcaccacc ggcaagctgc ccgtgccctg 550
gcccaccctc gtgaccaccc tgacctacgg cgtgcagtgc ttcagccgct 600
accccgacca catgaagcag cagcacttct tcaagtccgc catgcccga 650
ggctacgtcc aggagcgcac catcttcttc aaggacgacg gcaactacaa 700
gaccgcgccc gaggtgaagt tcgagggcga caccctggtg aaccgcatcg 750
agctgaaggg catcgacttc aaggaggacg gcaacatcct ggggcacaag 800
ctggagtaca actacaacag ccacaacgtc tatatcatgg ccgacaagca 850
gaagaacggc atcaaggtga acttcaagat ccgccacaac atcgaggacg 900
gcagcgtgca gctcgccgac cactaccagc agaacacccc catcggcgac 950
ggccccgtgc tgctgcccga caaccactac ctgagcaccc agtccgccct 1000
gagcaaagac cccaacgaga agcgcgatca catggtcctg ctggagttcg 1050
tgaccgccgc cgggatcact ctcggcattg acgagctgta caagaagctt 1100
agccatggct tcccgcggga ggtggaggag caggatgatg gcacgctgcc 1150

```

catgtcttgt gccaggaga gcgggatgga ccgtcaccct gcagcctgtg 1200  
cttctgctag gatcaatgtg tagtctagag tcggggcggc cgcccgcttc 1250  
gagcagacat gataagatac attgatgagt ttggacaaac cacaactaga 1300  
atgcagtga aaaaatgctt tatttgtgaa atttgtgatg ctattgcttt 1350  
atttghtaacc attataagct gcaataaaca agttaacaac aacaattgca 1400  
ttcattttat gtttcagggt cagggggagg tgtgggaggt tttttaagc 1450  
aagtaaaacc tctacaaatg tggtaaaatc gataaggatc cgtcgaccga 1500  
tgcccttgag agccttcaac ccagtcagct ccttccggtg ggccgggggc 1550  
atgactatcg tcgccgact tatgactgtc ttctttatca tgcaactcgt 1600  
aggacaggtg ccggcagcgc tcttccgctt cctcgctcac tgactcgctg 1650  
cgctcggtcg ttcggctgcg gcgagcggtg tcagctcact caaaggcggt 1700  
aatacgggta tccacagaat caggggataa cgcaggaaag aacatgtgag 1750  
caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg 1800  
tttttccata ggctccgcc ccctgacgag catcacaaa atcgacgctc 1850  
aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcggttc 1900  
cccctggaag ctccctcggt cgctctctcg ttcggaccct gccgcttacc 1950  
ggatacctgt ccgcctttct cccttcggga agcgtggcgc tttctcatag 2000  
ctcacgtgt aggtatctca gttcggtgta ggtcggtcgc tccaagctgg 2050  
gctgtgtgca cgaaccccc gttcagcccg accgctgcmc cttatccggt 2100  
aactatcgtc ttgagtccaa ccgggtaaga cagacttat cgccactggc 2150  
agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 2200  
cagagttctt gaagtgggtg cctaactacg gctacactag aaggacagta 2250  
tttggtatct gcgctctgct gaagccagtt accttcggaa aaagagttgg 2300  
tagctcttga tccggcaaac aaaccaccgc tggtagcggg ggtttttttg 2350  
tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct 2400  
ttgatctttt ctacgggggtc tgacgctcag tggaacgaaa actcacgtta 2450  
agggattttg gtcattgagat tatcaaaaag gatcttcacc tagatccttt 2500  
taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact 2550  
tggtctgaca gttaccaatg cttaatcagt gaggcaccta tctcagcgat 2600  
ctgtctattt cgttcatcca tagttgcctg actccccgtc gtgtagataa 2650  
ctacgatacg ggagggttta ccatctggcc ccagtgcctg aatgataccg 2700  
cgagaccac gctcaccggc tccagattta tcagcaataa accagccagc 2750  
cggaaggggc gagcgcagaa gtggctctgc aactttatcc gcctccatcc 2800  
agtctattaa ttggtgcccgg gaagctagag taagtagttc gccagttaat 2850  
agtttgcgca acgttggtgc cattgctaca ggcatcgtgg tgtcacgctc 2900  
gtcgtttggg atggcttcat tcagctccgg tttccaacga tcaaggcgag 2950  
ttacatgatc ccccatggtg tgcaaaaaag cggtagctc cttcggtcct 3000  
ccgatcggtg tcagaagtaa gttggccgca gtgttatcac tcatggttat 3050

```

ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 3100
ctgtgactgg tgagtactca accaagtcac tctgagaata gtgtatgceg 3150
cgaccgagtt gctcttgccc ggcgtcaata cgggataata ccgcgccaca 3200
tagcagaact ttaaaagtgc tcatcattgg aaaacgttct tcggggcgaa 3250
aactctcaag gatcttaccg ctgttgagat ccagttcgat gtaaccact 3300
cgtgcacca actgatcttc agcatctttt actttcacca gcgtttctgg 3350
gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga 3400
cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc 3450
atttatcagg gttattgtct catgagcgga tacatatgtg aatgtattta 3500
gaaaaataaa caaatagggg ttccgcgcac atttccccga aaagtgccac 3550
ctgacgcgcc ctgtagcggc gcattaagcg cggcggtgtg ggtgggttacg 3600
cgcagcgtga ccgctacact tgccagcgcc ctacgcgccg ctcttttcgc 3650
tttcttccct tcttttctcg ccacgttcgc cggttttccc cgtcaagctc 3700
taaategggg gctcccttta ggggtccgat ttagtgcttt acggcacctc 3750
gaccccaaaa aacttgatta gggatgatgg tcacgtagtg ggccatcgcc 3800
ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttctttaata 3850
gtggactctt gttccaaact ggaacaacac tcaaccctat ctcggtctat 3900
tcttttgatt tataagggat ttgcccatt tcggcctatt ggttaaaaaa 3950
tgagctgatt taacaaaaat ttaacgcgaa ttttaacaaa atattaacgt 4000
ttacaatttc ccattcgcca ttcaggctgc gcaactgttg ggaagggcga 4050
tcgggtgcggg cctcttcgct attacgccag cccaagctac catgataagt 4100
aagtaatat aaggtacggg aggtacttgg agcggccgca ataaaatatc 4150
tttattttca ttacatctgt gtgttggttt tttgtgtgaa tcgatagtac 4200
taacatacgc tctccatcaa aacaaaacga aacaaaacaa actagcaaaa 4250
taggctgtcc ccagtgcag tgcaggtgcc agaacatttc tctatcgata 4300

```

```

<210>      14
<211>      4987
<212>      DNA
<213>      artificial sequence
<220>
<223>      synthesized pNF-KB-Luc
<400>      14

```

```

ggtaccgagc tcttacgcgt gctagcggga atttccggga atttccggga 50
atttccggga atttccagat ctgccgcccc gactgcatct gcgtgttcga 100
attcgccaat gacaagacgc tgggcggggg ttgtgtcatc atagaactaa 150
agacatgcaa atatatttct tccggggaca ccgccagcaa acgcgagcaa 200

```

cgggccacgg	ggatgaagca	gaagcttggc	attccggtac	tgttggtaaa	250
gccaccatgg	aagacgccaa	aaacataaag	aaaggcccgg	cgccattcta	300
tccgctggaa	gatggaaccg	ctggagagca	actgcataag	gctatgaaga	350
gatacgccct	ggttcctgga	acaattgctt	ttacagatgc	acatatcgag	400
gtggacatca	cttacgctga	gtacttcgaa	atgtccgttc	ggttggcaga	450
agctatgaaa	cgatatgggc	tgaatacaaa	tcacagaatc	gtcgtatgca	500
gtgaaaactc	tcttcaattc	tttatgccgg	tgttgggcgc	gttattttatc	550
ggagttgcag	ttgcgcccgc	gaacgacatt	tataatgaac	gtgaattgct	600
caacagtatg	ggcatttcgc	agcctaccgt	ggtgttcggt	tccaaaaagg	650
ggttgcaaaa	aattttgaac	gtgcaaaaaa	agctcccaat	catccaaaaa	700
attattatca	tggattctaa	aacggattac	cagggatttc	agtcgatgta	750
cacgttcgtc	acatctcatc	tacctcccgg	ttttaatgaa	tacgattttg	800
tgccagagtc	cttcgatagg	gacaagacaa	ttgcactgat	catgaactcc	850
tctggatcta	ctggtctgcc	taaaggtgtc	gctctgcctc	atagaactgc	900
ctgcgtgaga	ttctcgcattg	ccagagatcc	tattttttggc	aatcaaatca	950
ttccggatac	tgcgatttta	agtgttggtc	cattccatca	cggtttttggg	1000
atgtttacta	cactcggata	tttgatatgt	ggatttcgag	tcgtcttaat	1050
gtatagattt	gaagaagagc	tgtttctgag	gagccttcag	gattacaaga	1100
ttcaaagtgc	gctgctgggtg	ccaaccctat	tctccttctt	cgccaaaagc	1150
actctgattg	acaaatacga	tttatctaat	ttacacgaaa	ttgcttctgg	1200
tggcgctccc	ctctctaagg	aagtcgggga	agcgggtgcc	aagagggtcc	1250
atctgccagg	tatcaggcaa	ggatatgggc	tcactgagac	tacatcagct	1300
attctgatta	cacccgaggg	ggatgataaa	ccgggcgcgg	tcggtaaagt	1350
tgttccattt	tttgaagcga	aggttgtgga	tctggatacc	gggaaaacgc	1400
tgggcgttaa	tcaaagaggc	gaactgtgtg	tgagaggtcc	tatgattatg	1450
tccggttatg	taaacaatcc	ggaagcgacc	aacgccttga	ttgacaagga	1500
tggatggcta	cattctggag	acatagctta	ctgggacgaa	gacgaacact	1550
tcttcatcgt	tgaccgcctg	aagtctctga	ttaaagtacaa	aggctatcag	1600
gtggctcccc	ctgaattgga	atccatcttg	ctccaacacc	ccaacatctt	1650
cgacgcaggt	gtcgcaggtc	ttcccgaagc	tgacgccggt	gaacttcccc	1700
ccgccgttgt	tgttttggag	cacggaaaga	cgatgacgga	aaaagagatc	1750
gtggattacg	tcgccagtca	agtaacaacc	gcgaaaaagt	tgcgcggagg	1800
agttgtgttt	gtggacgaag	taccgaaagg	tcttaccgga	aaactcgacg	1850
caagaaaaat	cagagagatc	ctcataaagg	ccaagaaggg	cggaaagatc	1900
gccgtgtaat	tctagagtcg	gggcggccgg	ccgcttcgag	cagacatgat	1950
aagatacatt	gatgagtttg	gacaaaccac	aactagaatg	cagtgaaaaa	2000
aatgccttat	ttgtgaaatt	tgtgatgcta	ttgctttatt	tgtaaccatt	2050
ataagctgca	ataaacaagt	taacaacaac	aattgcattc	attttatggt	2100



tcagggttcag ggggaggtgt gggaggtttt ttaaagcaag taaaacctct 2150  
acaaatgtgg taaaatcgat aaggatccgt cgaccgatgc ccttgagagc 2200  
cttcaaccca gtcagctcct tccggtgggc gcggggcatg actatcgtcg 2250  
ccgcacttat gactgtcttc tttatcatgc aactcgtagg acaggtgccg 2300  
gcagcgctct tccgcttcct cgctcactga ctcgctgcgc tcggtcgttc 2350  
ggctgcggcg agcggtatca gctcactcaa aggcggtaat acgggtatcc 2400  
acagaatcag gggataacgc aggaaagaac atgtgagcaa aaggccagca 2450  
aaaggccagg aaccgtaaaa aggccgcgtt gctggcgttt ttccataggc 2500  
tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg 2550  
cgaaacccga caggactata aagataccag gcgtttcccc ctggaagctc 2600  
cctcgtgcgc tctcctgttc cgaccctgcc gcttaccgga tacctgtccg 2650  
cctttctccc ttcggaagc gtggcgcttt ctcatagctc acgctgtagg 2700  
tatctcagtt cggtgtaggt cgctcgctcc aagctgggct gtgtgcacga 2750  
accccccggt cagcccgacc gctgcgcctt atccggtaac tatcgtcttg 2800  
agtccaaccc ggtaagacac gacttatcgc cactggcagc agccactggg 2850  
aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa 2900  
gtggtggcct aactacggct acactagaag gacagtattt ggtatctgcg 2950  
ctctgctgaa gccagttacc ttcggaaaaa gagttggtag ctcttgatcc 3000  
ggcaaacaaa ccaccgctgg tagcgggtgg ttttttgttt gcaagcagca 3050  
gattacgcgc agaaaaaaag gatctcaaga agatcctttg atcttttcta 3100  
cggggtctga cgctcagtgg aacgaaaact cacgttaagg gattttggtc 3150  
atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg 3200  
aagtttttaa tcaatctaaa gtatatatga gtaaaacttg tctgacagtt 3250  
accaatgctt aatcagttag gcacctatct cagcgatctg tctatttcgt 3300  
tcatccatag ttgcctgact ccccgctcgt tagataacta cgatacggga 3350  
gggcttacca tctggccccg gtgctgcaat gataccgcga gacccacgct 3400  
caccggctcc agatttatca gcaataaacc agccagccgg aagggccgag 3450  
cgcagaagtg gtcctgcaac tttatccgcc tccatccagt ctattaattg 3500  
ttgccgggaa gctagagtaa gtagttcgcc agttaatagt ttgcgcaacg 3550  
ttgttgccat tgctacaggc atcgtgggtg cagctcgtc gtttggtatg 3600  
gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc 3650  
catgttgtgc aaaaaagcgg ttagctcctt cggtcctccg atcgttgtca 3700  
gaagtaagtt ggccgcagtg ttatcactca tggttatggc agcactgcat 3750  
aattctctta ctgtcatgcc atccgtaaga tgcttttctg tgactggtga 3800  
gtactcaacc aagtcattct gagaatagtg tatgcggcga ccgagttgct 3850  
cttgcccggc gtcaatacgg gataataccg cgccacatag cagaacttta 3900  
aaagtgtca tcattggaaa acgttcttcg gggcgaaaac tctcaaggat 3950  
cttaccgctg ttgagatcca gttcgatgta acccactcgt gcacccaact 4000

gatcttcagc atcttttact ttcaccagcg tttctgggtg agcaaaaaca 4050  
ggaaggcaaa atgccgcaaa aaagggaata agggcgacac ggaaatgttg 4100  
aatactcata ctcttccttt ttcaatatta ttgaagcatt tatcagggtt 4150  
attgtctcat gagcggatac atatttgaat gtatttagaa aaataaacia 4200  
ataggggttc cgcgcacatt tccccgaaaa gtgccacctg acgcgccctg 4250  
tagcggcgca ttaagcgcgg cgggtgtggt ggttacgcgc agcgtgaccg 4300  
ctacacttgc cagcgcccta gcgcccgtc ctttcgcttt cttcccttcc 4350  
tttctcgcca cgcttcgocgg ctttccccgt caagctctaa atcggggggt 4400  
cccttttaggg ttccgattta gtgctttacg gcacctcgac ccaaaaaaac 4450  
ttgatttaggg tgatggttca cgtagtgggc catcgccctg atagacgggt 4500  
tttcgccctt tgacgttgga gtccacgttc tttaatagtg gactcttggt 4550  
ccaaactgga acaacactca accctatctc ggtctattct tttgatttat 4600  
aagggatttt gccgatttcg gcctattggg taaaaaatga gctgatttaa 4650  
caaaaattta acgcgaattt taacaaaata ttaacgttta caatttccca 4700  
ttcgccattc aggctgcgca actggtggga agggcgatcg gtgcgggcct 4750  
cttcgctatt acgccagccc aagctaccat gataagtaag taatattaag 4800  
gtacgggagg tacttggagc ggccgcaata aaatatcttt attttcatta 4850  
catctgtgtg ttggtttttt gtgtgaatcg atagtactaa catacgtct 4900  
ccatcaaaac aaaacgaaac aaaacaaact agcaaaatag gctgtcctca 4950  
gtgcaagtgc aggtgccaga acatttctct atcgata 4987